The GFS system for 2005 NIST SRE

GOLDEN FINGER SYSTEMS

Organization

- Introduction
- Implementation
  - Front end
  - Modeling
  - Segmentation
  - Scoring
- Results
- Conclusion

Introduction

- GFS – California based company
- Specializes in Biometrics Technologies
  - Fingerprint – Criminal AFIS and ID systems
  - Voice Recognition
  - Facial Systems
- Works with Affiliates in both underlying technologies and applications

Introduction

- The second year of participating in NIST SRE
- Participant tasks:
  - 1 conv4w training – 1 conv4w testing
  - 1 conv4w training – 1 conv2w testing
- Standard UBM-GMM system based on short-term acoustic information

Implementation

Front end

- Band limits (300Hz ~ 3400 Hz)
- Energy based SAD
- 20 ms frame length, 10 ms frame shift
- 20 MFCCs from 21 filters
- 20 Deltas appended
- CMS and CVN
- Feature mapping

Implementation

Modeling

- UBM
  - Gender-dependent
  - 1024 component diagonal GMMs
  - Data from NIST SRE 97 2001 2003 2004 database, about 30 hours for each gender
- Channel UBM
  - 5 channel UBMs for each gender: CDMA, GSM, Cordless, Elec, Carbon.
  - Built for corresponding UBM by mean and variance adaptation

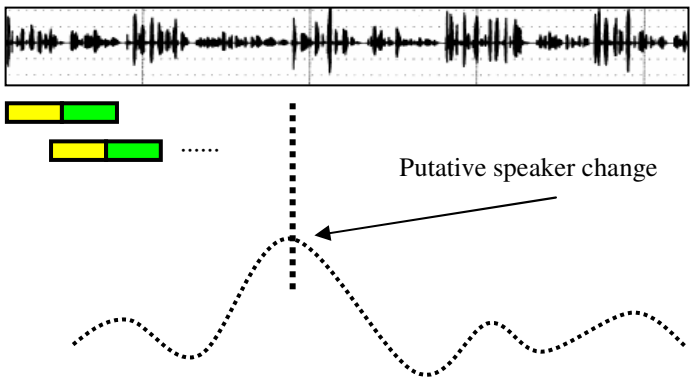
Implementation

Modeling

- Speaker Model
  - Built by Mean-only adaptation from corresponding gender UBM
  - Relevance factor set to 16
- T-norm Model
  - 100 female speakers and 100 male speakers from NIST 2001 Evaluation data
  - Built by Mean-only adaptation from corresponding gender UBM

Implementation
Segmentation

- Speaker change detection
  - Use DIST-BIC based change detection*
  - Implementation:



- Compute GLR distance between sliding adjacent windows to produce a distance curve
- The local maxima on the curve is the putative speaker change when the differences between its value and those of the minima surrounding it are above a certain threshold

"DISTBIC:A speaker-based segmentation for audio data indexing", P. Delacourt *, C.J. Wellekens, Speech Communication 32 (2000) 111~126

Implementation
Segmentation

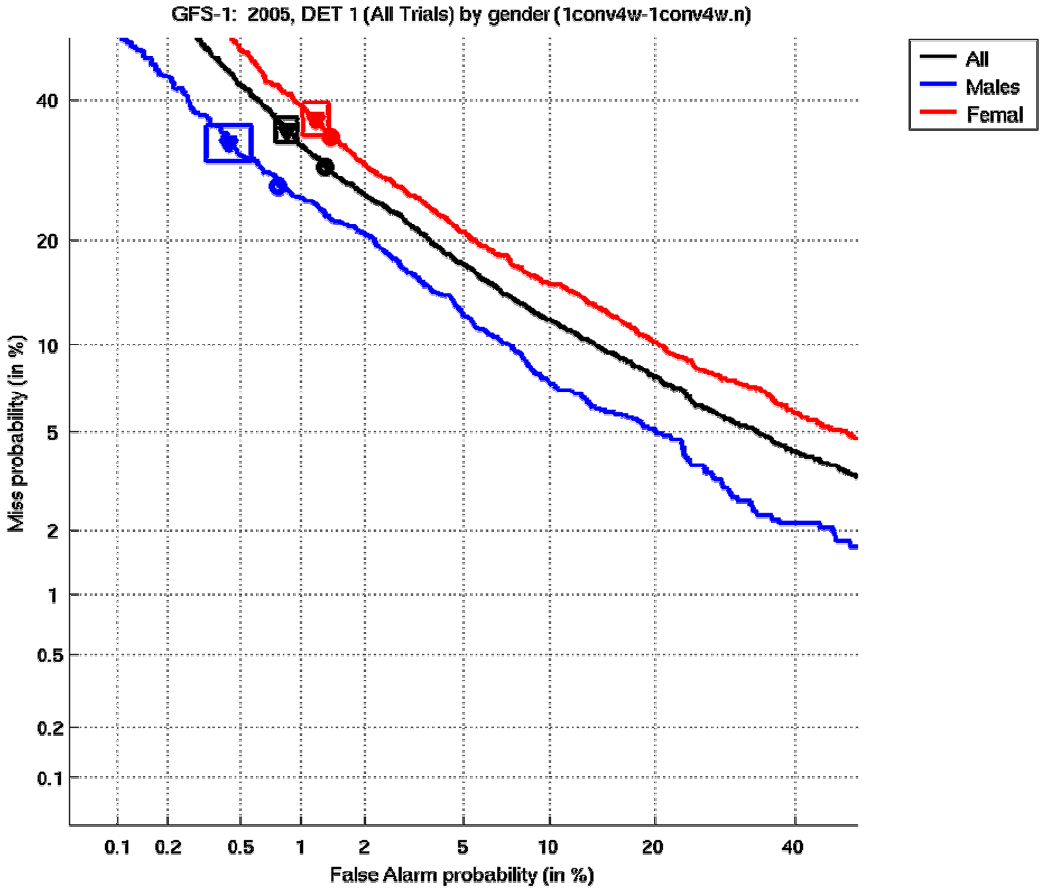
- Clustering
  - Initialize the clusters with segments from speaker change detection
  - Perform agglomerative clustering using GLR until we get 2 clusters for each 2-sp speech
- Refining
  - Construct GMMs for the 2 clusters and reassign frames to models

Scoring

- For each test segment, Log-likelihood ratio is computed against target GMM and corresponding gender UBM
- Only scoring the 3 top Gaussians
- T-Norm is performed to normalize the score
- The threshold is set to reach the minimum DCF in NIST 2004 SRE

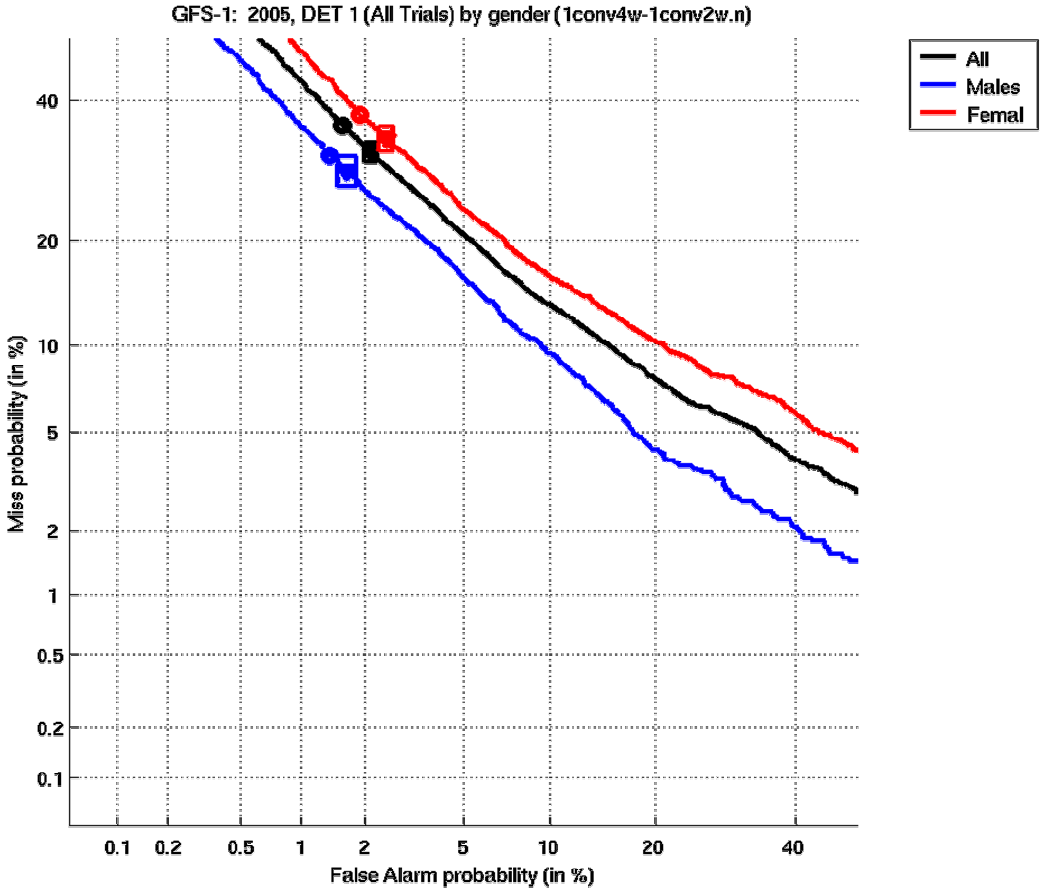
Results

1 conv4w training – 1 conv4w testing



Results

1 conv4w training – 1 conv2w testing



Results

- Post evaluation analysis
  - The performance on male speakers is much better than on female, approximately 7% at the EER.
  - Some errors might be brought by feature mapping:
    - we found that 15% of all data for training channel UBMs is wrongly classified when feature mapping is applied in this system
  - Significant performance loss in 2-sp condition due to segmentation errors

Conclusion

- Standard UBM-GMM system for 2005 NIST SRE
- Future work will focus on:
  - Error analysis
  - More robust channel compensation
  - Combine baseline system with other systems using high-level information